How many Laplace transforms of probability measures are there?

Fuchang Gao* Wenbo V. Li† Jon A. Wellner‡

January 1, 2010

Abstract

A bracketing metric entropy bound for the class of Laplace transforms of probability measures on $[0, \infty)$ is obtained through its connection with the small deviation probability of a smooth Gaussian process. Our results for the particular smooth Gaussian process seem to be of independent interest.

Keywords: Laplace Transform; bracketing metric entropy; completely monotone functions; smooth Gaussian process; small deviation probability

*Department of mathematics, University of Idaho, fuchang@uidaho.edu.

†Department of Mathematical Sciences, University of Delaware, wli@math.udel.edu. Supported in part by NSF grant DMS-0805929.

‡Department of Statistics, University of Washington, jaw@stat.washington.edu. Supported in part by NSF Grant DMS-0804587

1 Introduction

Let μ be a finite measure on $[0, \infty)$. The Laplace transform of ν is a function on $(0, \infty)$ defined by

$$f(t) = \int_0^\infty e^{-ty} \mu(dy). \tag{1}$$

It is easy to check that such a function has the property that $(-1)^n f^{(n)}(t) \geq 0$ for all non-negative integer n and all $t > 0$. A function on $(0, \infty)$ with this property is called a completely monotone function on $(0, \infty)$. A characterization due to Bernstein (c.f. Williamson (1956)) says that f is completely monotone on $(0, \infty)$ if and only if there is a non-negative measure μ (not necessary finite) on $[0, \infty)$ such that (1) holds. Therefore, due to monotonicity, the class of Laplace transforms of finite measures on $[0, \infty)$ is the same as the class of bounded completely monotone functions on $(0, \infty)$. These functions can be extended to continuous functions on $[0, \infty)$, and we will call them completely monotone on $[0, \infty)$.

Completely monotonic functions have remarkable applications in various fields, such as probability and statistics, physics and potential theory. The main properties of these functions are given in Widder (1941), Chapter IV. For example, the class of completely monotonic functions is closed under sums, products and pointwise convergence. We refer to Alzer and Berg (2002) for a detailed list of references on completely monotonic functions. Closely related to the class of completely monotonic functions are the so-called k-monotone functions, where the non-negativity of $(-1)^n f^{(n)}$ is required for all integers $n \leq k$. In fact, completely monotonic functions can be viewed as the limiting case of the k-monotone functions as $k \to \infty$. In this sense, the present work is a partial extension of Gao (2008) and Gao and Wellner (2009).

Let $\mathcal{M}_\infty$ be the class of completely monotone functions on $[0, \infty)$ that are bounded by 1. Then

$$\mathcal{M}_\infty = \left\{ f : [0, \infty) \to [0, \infty) \middle| f(t) = \int_0^\infty e^{-tx} \mu(dx), \|\mu\| \leq 1 \right\}.$$

It is well known (see e.g. Feller (1971), Theorem 1, page 439) that the sub-class of $\mathcal{M}_\infty$ with $f(0) = 1$ corresponds exactly to the Laplace transforms of the class of probability measures μ on $[0, \infty)$. For a random variable with distribution function $F(t) = P(X \leq t)$, the *survival function* $S(t) = 1 - F(t) = P(X > t)$. Thus the class

$$\mathcal{S}_\infty = \left\{ S : [0, \infty) \to [0, \infty) \middle| S(t) = \int_0^\infty e^{-tx} \mu(dx), \|\mu\| = 1 \right\}$$

is exactly the class of survival functions of all scale mixtures of the standard exponential distribution (with survival function e^{-t}), with corresponding densities

$$p(t) = -S'(t) = \int_0^\infty x e^{-xt} \mu(dx), \qquad t \geq 0.$$

It is easily seen that the class $\mathcal{P}_\infty$ of such densities with $p(0) < \infty$ is also a class of completely monotone functions corresponding to probability measures μ on $[0, \infty)$ with finite first moment. These classes have many applications in statistics; see e.g. Jewell (1982) for a brief survey. Jewell (1982) considered nonparametric estimation of a completely monotone density and showed that the nonparametric maximum likelihood estimator (or MLE) for this class is almost surely consistent. The bracketing entropy bounds derived below can be considered as a first step toward global rates of convergence of the MLE.

In probability and statistical applications, one way to understand the complexity of a function class is by way of the metric entropy for the class under certain common distances. Recall that the metric entropy of a function class $\mathcal{F}$ under distance ρ is defined to be $\log N(\varepsilon, \mathcal{F}, \rho)$ where $N(\varepsilon, \mathcal{F}, \rho)$ is the minimum number of open balls of radius ε needed to cover $\mathcal{F}$. In statistical applications, sometimes we also need bracketing metric entropy which is defined as $\log N_{[]}(\varepsilon, \mathcal{F}, \rho)$ where

$$N_{[]}(\varepsilon, \mathcal{F}, \rho) := \min \left\{ n : \exists \underline{f}_1, \overline{f}_1, \ldots, \underline{f}_n, \overline{f}_n \text{ s.t. } \rho(\overline{f}_k, \underline{f}_k) \le \varepsilon, \mathcal{F} \subset \bigcup_{k=1}^{n} [\underline{f}_k, \overline{f}_k] \right\},$$

and

$$[\underline{f}_k, \overline{f}_k] = \left\{ g \in \mathcal{F} : \underline{f}_k \le g \le \overline{f}_k \right\}.$$

Clearly $N(\varepsilon, \mathcal{F}, \rho) \le N_{[]}(\varepsilon, \mathcal{F}, \rho)$ and they are close related in our setting below.

In this paper, we study the metric entropy of $\mathcal{M}_\infty$ under the $L^p(\nu)$ norm given by

$$\|f\|^p_{L^p(\nu)} = \int_0^\infty |f(x)|^p \nu(dx), \quad 1 \le p \le \infty,$$

where ν is a probability measure on $[0, \infty)$. Our main result is the following

Theorem 1.1. (i) Let ν be a probability measure on $[0, \infty)$. There exists a constant C depending only on $p \ge 1$ such that for any $0 < \varepsilon < 1/4$,

$$\log N_{[]}(\varepsilon, \mathcal{M}_\infty, \|\cdot\|_{L^p(\nu)}) \le C \log(\Gamma/\gamma) \cdot |\log \varepsilon|^2,$$

for any $0 < \gamma < \Gamma < \infty$ such that $\nu([\gamma, \Gamma]) \ge 1 - 4^{-p}\varepsilon^p$. In particular, if there exists a constant $K > 1$, such that $\nu([\varepsilon^K, \varepsilon^{-K}]) \ge 1 - 4^{-p}\varepsilon^p$, then

$$\log N_{[]}(\varepsilon, \mathcal{M}_\infty, \|\cdot\|_{L^p(\nu)}) \le CK |\log \varepsilon|^3.$$

(ii) If ν is Lebesgue measure on $[0, 1]$, then

$$\log N_{[]}(\varepsilon, \mathcal{M}_\infty, \|\cdot\|_{L^2(\nu)}) \asymp \log N(\varepsilon, \mathcal{M}_\infty, \|\cdot\|_{L^2(\nu)}) \asymp |\log \varepsilon|^3,$$

where $A \asymp B$ means there exist universal constants $C_1, C_2 > 0$ such that $C_1 A \le B \le C_2 B$.

As an equivalent result for part (ii) of the above theorem, we have the following important small deviation probability estimates for an associated smooth Gaussian process. In particular, it may be of interest to find a probabilistic proof for the lower bound directly.

Theorem 1.2. Let $Y(t)$, $t > 0$, be a Gaussian process with covariance $\mathbb{E}\,Y(t)Y(s) = (1-e^{-t-s})/(t+s)$, then for $0 < \varepsilon < 1$

$$\log \mathbb{P}\left(\sup_{t>0} |Y(t)| < \varepsilon\right) \asymp -|\log \varepsilon|^3.$$

The rest of the paper is organized as follows. In Section 2, we provide the upper bound estimate in the main result by explicit construction. In Section 3, we summarize various connections between entropy numbers of a set (and its convex hull) and small ball probability for the associated Gaussian process. Some of our observations in a general setting are stated explicitly for the first time. Finally we identify the particular Gaussian process suitable for our entropy estimates. Then in Section 4, we obtain the required upper bound small ball probability estimate (which implies the lower bound entropy estimates as discussed in section 3) by a simple determinant estimates. This method of small ball estimates is made explicit here for the first time and can be used in many more problems. The technical determinant estimates are also of independent interests.

2 Upper Bound Estimate

In this section, we provide an upper bound for $N_{[\,]}(\varepsilon, \mathcal{M}_\infty, \|\cdot\|_{L^p(\nu)})$, where ν is a probability measure on $[0,\infty)$ and $1 \le p \le \infty$. This is accomplished by an explicit construction of ε-brackets under $L^p(\nu)$ distance.

For each $0 < \varepsilon < 1/4$, we choose $\gamma > 0$ and $\Gamma = 2^m\gamma$ where m is a positive integer such that $\nu([\gamma, \Gamma]) \ge 1 - 4^{-p}\varepsilon^p$. We use the notion $\mathbb{I}(a \le t < b)$ to denote the indicator function of the interval $[a, b)$. Now for each $f \in \mathcal{M}_\infty$, we first write in block form

$$f(t) = \mathbb{I}(0 \le t < \gamma)f(t) + \mathbb{I}(t \ge \Gamma)f(t) + \sum_{i=1}^{m} \mathbb{I}(2^{i-1}\gamma \le t < 2^i\gamma)f(t).$$

Then for each block $2^{i-1}\gamma \le t < 2^i\gamma$, we separate the integration limits at the level $2^{2-i}|\log\varepsilon|/\gamma$ and use the first N terms of Taylor's series expansion of e^{-u} with error terms associated with $\xi = \xi_{u,N}$, $0 \le \xi \le 1$, to rewrite

$$f(t) = \mathbb{I}(0 \le t < \gamma)f(t) + \mathbb{I}(t \ge \Gamma)f(t) + \sum_{i=1}^{m}(p_i(t) + q_i(t) + r_i(t))$$

where

$$\begin{aligned}
p_i(t) &=: \mathbb{I}(2^{i-1}\gamma \le t < 2^i\gamma)\sum_{n=0}^{N}\frac{(-1)^n t^n}{n!}\int_0^{2^{2-i}|\log\varepsilon|/\gamma} x^n \mu(dx)\\
q_i(t) &=: \mathbb{I}(2^{i-1}\gamma \le t < 2^i\gamma)\int_0^{2^{2-i}|\log\varepsilon|/\gamma} \frac{(-\xi t x)^{N+1}}{(N+1)!}\mu(dx)\\
r_i(t) &=: \mathbb{I}(2^{i-1}\gamma \le t < 2^i\gamma)\int_{2^{2-i}|\log\varepsilon|/\gamma}^{\infty} e^{-tx}\mu(dx).
\end{aligned}$$

We choose the integer N so that

$$4e^2|\log\varepsilon| - 1 \le N < 4e^2|\log\varepsilon|. \tag{2}$$

Then, by using the inequality $k! \ge (k/e)^k$ and the fact that $0<\xi<1$, we have within the block $2^{i-1}\gamma \le t < 2^i\gamma$,

$$\begin{aligned} |q_i(t)| &\le \int_0^{2^{2-i}|\log\varepsilon|/\gamma} \frac{(tx)^{N+1}}{(N+1)!}\mu(dx) \\ &\le \frac{|4\log\varepsilon|^{N+1}}{(N+1)!} \le \left(\frac{4e|\log\varepsilon|}{N+1}\right)^{N+1} \le e^{-(N+1)} \le \varepsilon^{4e^2}, \end{aligned}$$

where we used $tx \le 2^i\gamma\cdot 2^{2-i}|\log\varepsilon|/\gamma = 4|\log\varepsilon|$ in the second inequality above. This implies, due to disjoint supports of $q_i(t)$,

$$\left|\sum_{i=1}^m q_i(t)\right| \le \varepsilon^{4e^2}. \tag{3}$$

Next, we notice that for $t \ge 2^{i-1}\gamma$ and $x \ge 2^{2-i}\gamma^{-1}|\log\varepsilon|$, $e^{-tx} \le \varepsilon^2$. Thus

$$\left|\sum_{i=1}^m r_i(t)\right| \le \sum_{i=1}^m \mathbb{I}(2^{i-1}\gamma \le t < 2^i\gamma)\int_{2^{2-i}\gamma^{-1}|\log\varepsilon|}^\infty \varepsilon^2\mu(dx) \le \varepsilon^2. \tag{4}$$

Finally, because $|f|\le 1$ and $\nu([0,\gamma)) + \nu([\Gamma,\infty)) \le 4^{-p}\varepsilon^p$, we have

$$\|1_{0\le t<\gamma}f(t) + 1_{t\ge\Gamma}f(t)\|_{L^p(\nu)} \le \varepsilon/4.$$

Together with (3) and (4), we see that the set

$$\mathcal{R} =: \left\{\sum_{i=1}^m q_i(t) + \sum_{i=1}^m r_i(t) + \mathbb{I}(t<\gamma)f(t) + \mathbb{I}(t\ge\Gamma)f(t) : f \in \mathcal{M}_\infty\right\}$$

has diameter in $L^p(\nu)$-distance at most $\varepsilon^2 + \varepsilon^{4e^2} + \varepsilon/4 < \varepsilon/2$.

Therefore, if we denote $\mathcal{P}_i = \{p_i(t) : f \in \mathcal{M}_\infty\}$, then the expansion of f above implies that $\mathcal{M}_\infty \subset \sum_{i=1}^m \mathcal{P}_i + \mathcal{R}$, and consequently, we have

$$N_{[]}(\varepsilon, \mathcal{M}_\infty, \|\cdot\|_{L^p(\nu)}) \le N_{[]}\left(\varepsilon/2, \sum_{i=1}^m \mathcal{P}_i, \|\cdot\|_{L^p(\nu)}\right).$$

For any $1\le i\le m$ and any $p_i \in \mathcal{P}_i$, we can write

$$p_i(t) = \mathbb{I}(2^{i-1}\gamma \le t < 2^i\gamma)\sum_{n=0}^{N_i}(-1)^n a_{ni}(2^{-i}\gamma^{-1}t)^n, \tag{5}$$

where $0 \le a_{ni} \le |4\log\varepsilon|^n/n!$. Now we can construct

$$\overline{p}_i = \mathbb{I}(2^{i-1}\gamma \le t < 2^i\gamma)\sum_{n=0}^{N}(-1)^n b_{ni}(2^{-i}\gamma^{-1}t)^n,$$
$$\underline{p}_i = \mathbb{I}(2^{i-1}\gamma \le t < 2^i\gamma)\sum_{n=0}^{N}(-1)^n c_{ni}(2^{-i}\gamma^{-1}t)^n,$$

where

$$b_{ni} = \begin{cases} \frac{\varepsilon}{2^{n+2}}\left\lceil \frac{2^{n+2}a_{ni}}{\varepsilon}\right\rceil & n \text{ is even} \\ \frac{\varepsilon}{2^{n+2}}\left\lfloor \frac{2^{n+2}a_{ni}}{\varepsilon}\right\rfloor & n \text{ is odd} \end{cases}, \quad c_{ni} = \begin{cases} \frac{\varepsilon}{2^{n+2}}\left\lfloor \frac{2^{n+2}a_{ni}}{\varepsilon}\right\rfloor & n \text{ is even} \\ \frac{\varepsilon}{2^{n+2}}\left\lceil \frac{2^{n+2}a_{ni}}{\varepsilon}\right\rceil & n \text{ is odd} \end{cases}.$$

Clearly, $\underline{p}_i(t) \le p_i(t) \le \overline{p}_i(t)$, and

$$\begin{aligned} |\overline{p}_i - \underline{p}_i| &\le \mathbb{I}(2^{i-1}\gamma \le t < 2^i\gamma)\sum_{n=0}^{N}|c_{ni}-b_{ni}|(2^{-i}\gamma^{-1}t)^n \\ &\le \mathbb{I}(2^{i-1}\gamma \le t < 2^i\gamma)\sum_{n=0}^{N}\frac{\varepsilon}{2^{n+2}}(2^{-i}\gamma^{-1}t)^n \\ &\le \frac{\varepsilon}{2}\mathbb{I}(2^{i-1}\gamma \le t < 2^i\gamma). \end{aligned}$$

Hence

$$\sum_{i=1}^m \underline{p}_i \le \sum_{i=1}^m p_i \le \sum_{i=1}^m \overline{p}_i \le \sum_{i=1}^m \underline{p}_i + \varepsilon/2.$$

That is, the sets

$$\underline{\mathcal{P}} =: \left\{\sum_{i=1}^m \underline{p}_i : p_i \in \mathcal{P}_i, 1 \le i \le m\right\} \quad \text{and} \quad \overline{\mathcal{P}} =: \left\{\sum_{i=1}^m \overline{p}_i : p_i \in \mathcal{P}_i, 1 \le i \le m\right\}$$

form $\varepsilon/2$ brackets of $\sum_{i=1}^m \mathcal{P}_i$ in L^∞-norm, and thus in $L^p(\nu)$-norm for all $1 \le p < \infty$.

Now we count the number of different realizations of $\overline{\mathcal{P}}$ and $\underline{\mathcal{P}}$. Note that, due to the uniform bound on a_{ni} in (5) there are no more than

$$\frac{2^{n+1}}{\varepsilon}\cdot\frac{|4\log\varepsilon|^n}{n!} + 1$$

realizations for b_{ni}. So, the number of realizations of $\overline{p}_i$ is bounded by

$$\prod_{n=0}^{N}\left(\frac{2^{n+1}}{\varepsilon}\cdot\frac{|4\log\varepsilon|^n}{n!} + 1\right).$$

Because $n! > (n/e)^n$, for all $1 \le n \le N$, we have

$$\frac{2^{n+1}}{\varepsilon}\cdot\frac{|4\log\varepsilon|^n}{n!} + 1 \le \frac{3}{\varepsilon}\left(\frac{8e|\log\varepsilon|}{n}\right)^n.$$

Thus, the number of realizations of $\overline{p}_i$ is bounded by

$$
\begin{aligned}
& \left(\frac{3}{\varepsilon}\right)^{N+1} \cdot \exp\left(\sum_{n=1}^{N}(n \log|8e \log \varepsilon| - n \log n)\right) \\
\leq \quad & \left(\frac{3}{\varepsilon}\right)^{N+1} \cdot \exp\left(\frac{N(N+1)}{2} \log|8e \log \varepsilon| - \int_1^N x \log x dx\right) \\
\leq \quad & \left(\frac{3}{\varepsilon}\right)^{N+1} \cdot \exp\left(\frac{N(N+1)}{2} \log|8e \log \varepsilon| - \frac{N^2}{2} \log N + \frac{N^2}{4}\right) \\
\leq \quad & \exp\left(C|\log \varepsilon|^2\right)
\end{aligned}
$$

for some absolute constant C, where in the last inequality we used the bounds on N given in (2).

Hence the total number of realizations of $\overline{\mathcal{P}}$ is bounded by $\exp(Cm|\log \varepsilon|^2)$. Similar estimate holds for the total number of realizations of $\underline{\mathcal{P}}$, and we finally obtain

$$\log N_{[]}(\varepsilon, \mathcal{M}_\infty, \|\cdot\|_{L^p(\nu)}) \leq C' m |\log \varepsilon|^2$$

for some different constant C'. This finishes the proof since $m = \log_2(\Gamma/\gamma)$.

3 Entropy of Convex Hulls

A lower bound estimate of metric entropy is typically difficult, because it often involves a construction of a well-separated set of maximal cardinality. Thus we introduce some soft analytic arguments to avoid this difficulty and change the problem into a familiar one in this section. The hard estimates are given in the next section.

First note that $\mathcal{M}_\infty$ is just the convex hull of the functions $k_s(\cdot)$, $0 < s < \infty$, where $k_s(t) = e^{-ts}$. We recall a general method about the entropy of convex hulls that was introduced in Gao (2004). Let T be a set in $\mathbb{R}^n$ or in a Hilbert space. The convex hull of T can be expressed as

$$\operatorname{conv}(T) = \left\{\sum_{n=1}^{\infty} a_n t_n : t_n \in T, a_n \geq 0, n \in \mathbb{N}, \sum_{n=1}^{\infty} a_n = 1\right\};$$

while the absolute convex hull of T is defined by

$$\operatorname{abconv}(T) = \left\{\sum_{n=1}^{\infty} a_n t_n : t_n \in T, n \in \mathbb{N}, \sum_{n=1}^{\infty} |a_n| \leq 1\right\}.$$

Clearly, by using probability measures and signed measures, we can express

$$
\begin{aligned}
\operatorname{conv}(T) &= \left\{\int_T t\mu(dt) : \mu \text{ is a probability measure on } T\right\}; \\
\operatorname{abconv}(T) &= \left\{\int_T t\mu(dt) : \mu \text{ is a signed measure on } T, \|\mu\|_{TV} \leq 1\right\}.
\end{aligned}
$$

For any norm $\|\cdot\|$ the following is clear:

$$\mathrm{conv}(T) \subset \mathrm{abconv}(T) \subset \mathrm{conv}(T) - \mathrm{conv}(T).$$

Therefore,

$$N(\varepsilon, \mathrm{conv}(T), \|\cdot\|) \le N(\varepsilon, \mathrm{abconv}(T), \|\cdot\|) \le [N(\varepsilon/2, \mathrm{conv}(T), \|\cdot\|)]^2.$$

In particular, at the logarithmic level, the two entropy numbers are comparable, modulo constant factors on ε. The benefit of using absolute convex hull is that it is symmetric and can be viewed as the unit ball of a Banach space, which allows us to use the following duality lemma of metric entropy:

$$\log N(\varepsilon, \mathrm{abconv}(T), \|\cdot\|) \asymp \log N(c_2\varepsilon, B, \|\cdot\|_T)$$

where B is dual ball of the norm $\|\cdot\|$, and $\|\cdot\|_T$ is the norm introduced by T, that is,

$$\|x\|_T := \sup_{t\in T} |\langle t, x\rangle| = \sup_{t\in \mathrm{abconv}(T)} |\langle t, x\rangle|.$$

Strictly speaking, the duality lemma remains as a conjecture in the general case. However, when the norm $\|\cdot\|$ is the Hilbert space norm, this has been proved. See Tomczak-Jaegermann (1987), Bourgain et al. (1989), and Artstein et al. (2004).

A striking relation discovered by Kuelbs and Li (1993) says that the entropy number $\log N(\varepsilon, B, \|\cdot\|_T)$ is determined by the Gaussian measure of the set

$$D_\varepsilon =: \{x \in H : \|x\|_T \le \varepsilon\}$$

under some very weak regularity assumptions. For details, see Kuelbs and Li (1993), Li and Linde (1999), and also Corollary 2.2 of Aurzada et al. (2008). Using this relation, we can now summarize the connection between metric entropy of convex hulls and Gaussian measure of D_ε into the following

Proposition 3.1. Let T be a precompact set in a Hilbert space. For $\alpha > 0$ and $\beta \in \mathbb{R}$,

$$\log \mathbb{P}(D_\varepsilon) \le -C_1\varepsilon^{-\alpha}|\log\varepsilon|^\beta$$

if and only if

$$\log N(\varepsilon, \mathrm{conv}(T), \|\cdot\|) \ge C_2\varepsilon^{-\frac{2\alpha}{2+\alpha}}|\log\varepsilon|^{\frac{2\beta}{2+\alpha}};$$

and for $\beta > 0$ and $\gamma \in \mathbb{R}$,

$$\log \mathbb{P}(D_\varepsilon) \le -C_1|\log\varepsilon|^\beta(\log|\log\varepsilon|)^\gamma$$

if and only if

$$\log N(\varepsilon, \mathrm{conv}(T), \|\cdot\|_2) \ge C_2|\log\varepsilon|^\beta(\log|\log\varepsilon|)^\gamma.$$

Furthermore, the results also hold if the directions of the inequalities are switched.

The result of this proposition can be implicitly seen in Gao (2004), where an explanation of the relation between $N(\varepsilon, B, \|\cdot\|_T)$ and the Gaussian measure of D_ε is also given.

Perhaps, the most useful case of Proposition 3.1 is when T is a set of functions: $K(t,\cdot)$, $t \in T$, where for each fixed $t \in T$, $K(t,\cdot)$ is a function in $L^2(\Omega)$, and where Ω is a bounded set in $\mathbb{R}^d$, $d \geq 1$. For this special case, we have

Corollary 3.2. Let $X(t) = \int_\Omega K(t,x)dB(x)$, $t \in T$, where $K(t,\cdot)$ are square-integrable functions on a bounded set Ω in $\mathbb{R}^d$, $d \geq 1$, and $B(x)$ is the d-dimensional Brownian sheet on Ω. If $\mathcal{F}$ is the convex hull of the functions $K(\cdot,\omega)$, $\omega \in \Omega$, then

$$\log \mathbb{P}\left(\sup_{t\in T} |X(t)| < \varepsilon\right) \asymp \varepsilon^{-\alpha} |\log \varepsilon|^\beta$$

for $\alpha > 0$ and $\beta \in \mathbb{R}$ if and only if

$$\log N(\varepsilon, \mathcal{F}, \|\cdot\|) \asymp \varepsilon^{-\frac{2\alpha}{2+\alpha}} |\log \varepsilon|^{\frac{2\beta}{2+\alpha}};$$

and for $\beta > 0$ and $\gamma \in \mathbb{R}$,

$$\log \mathbb{P}\left(\sup_{t\in T} |X(t)| < \varepsilon\right) \asymp -|\log \varepsilon|^\beta (\log|\log \varepsilon|)^\gamma$$

if and only if

$$\log N(\varepsilon, \mathcal{F}, \|\cdot\|_2) \asymp |\log \varepsilon|^\beta (\log|\log \varepsilon|)^\gamma.$$

The authors found this corollary especially useful. For example, it was used in Blei et al. (2007) and Gao (2008) to change a problem of metric entropy to a problem of small deviation probability of a problem about a Gaussian process which is relatively easier. The proof is given in Gao (2008) for the case $\Omega = [0,1]$, and in Blei et al. (2007) for the case $[0,1]^d$. For the general case, it can be proved as easily. Indeed, the only thing we need to prove is that $\mathbb{P}(D_\varepsilon)$ can be expressed as the probability of the set $\sup_{t\in T} |X(t)| < \varepsilon$. We outline a proof below. Let ϕ_n be an orthonormal basis of $L^2(\Omega)$, then

$$X(t) = \int_\Omega K(t,s)dB(s) = \sum_{n=1}^\infty \xi_n \int_\Omega K(t,s)\phi_n(s)ds$$

where ξ_n are i.i.d standard normal random variables. Thus,

$$\begin{aligned}
\mathbb{P}(D_\varepsilon) &= \mathbb{P}\left\{g \in L^2(\Omega) : \left|\int_\Omega f(s)g(s)ds\right| < \varepsilon, f \in \mathcal{F}\right\} \\
&= \mathbb{P}\left\{g \in L^2(\Omega) : \left|\int_T \int_\Omega K(t,s)g(s)ds\mu(dt)\right| < \varepsilon, \|\mu\|_{TV} \leq 1\right\} \\
&= \mathbb{P}\left\{\sum_{n=1}^\infty a_n\phi_n(s) : \sum_{n=1}^\infty a_n^2 < \infty, \left|\sum_{n=1}^\infty a_n \int_T \int_\Omega K(t,s)\phi_n(s)ds\mu(dt)\right| < \varepsilon, \|\mu\|_{TV} \leq 1\right\} \\
&= \mathbb{P}\left\{\sum_{n=1}^\infty a_n\phi_n(s) : \sum_{n=1}^\infty a_n^2 < \infty, \sup_{t\in T}\left|\sum_{n=1}^\infty a_n \int_\Omega K(t,s)\phi_n(s)ds\right| < \varepsilon\right\} \\
&= \mathbb{P}\left(\sup_{t\in T} |X(t)| < \varepsilon\right).
\end{aligned}$$

Now back to our problem of estimate $\log N(\varepsilon, \mathcal{M}_\infty, \|\cdot\|_2)$ in the statement of (ii) of the theorem, where $\|\cdot\|_2$ is the L^2 norm under the Lebesgue measure on $[0,1]$, we notice that $\mathcal{M}_\infty$ is the convex hull of the functions $C(\cdot, s)$, $s \in [0,\infty)$, on $[0,1]$ with $C(t,s) = e^{-ts}$. However, $[0,\infty)$ is not bounded. In order to use Corollary 3.2, we need to make a change of variables. Notice that by letting $y = e^{-s}$, we can view $\mathcal{M}_\infty$ as convex hull of $K(\cdot, y)$, $y \in (0,1]$, where $K(t,y) = y^t$. Clearly, $K(t,\cdot)$ are square-integrable functions on the bounded set $(0,1]$. Now, for this K, the corresponding $X(t)$ is a Gaussian process on $[0,1]$ with covariance

$$\mathbb{E}\, X(t)X(s) = \frac{ts-1}{\log(ts)}, \quad s,t \in (0,1], \quad (s,t) \neq (1,1),$$

and $\mathbb{E}\, X(1)^2 = 1$. Thus, the problem becomes how to estimate

$$\mathbb{P}\left(\sup_{t\in(0,1]} |X(t)| < \varepsilon\right), \text{ or equivalently } \mathbb{P}\left(\sup_{t\geq 0} |Y(t)| < \varepsilon\right),$$

where $Y(t) = X(e^{-t})$, which has covariance structure

$$\mathbb{E}\, Y(t)Y(s) = \frac{1-e^{-t-s}}{t+s}, \quad s,t \geq 0. \tag{6}$$

We now turn to the lower estimates of this probability.

4 Lower Bound Estimate

Let $Y(t)$, $t \geq 0$ be the centered Gaussian process defined in (6). Our goal in this section is to prove that

$$\log \mathbb{P}(\sup_{t\geq 0} |Y(t)| < \varepsilon) \leq -C|\log \varepsilon|^3,$$

for some constant $C > 0$.

Note that for any sequence of positive numbers $\{\delta_i\}_{i=1}^n$,

$$\begin{aligned}
\mathbb{P}\left(\sup_{t\geq 0} |Y(t)| < \varepsilon\right) &\leq \mathbb{P}(\max_{1\leq i\leq n} |Y(\delta_i)| < \varepsilon) \\
&= (2\pi)^{-n/2}(\det\Sigma)^{-1/2} \int_{\max_{1\leq i\leq n}|y_i|\leq\varepsilon} \exp\left(-\langle y, \Sigma^{-1}y\rangle\right) dy_1\cdots dy_n \\
&\leq (2\pi)^{-n/2}(\det\Sigma)^{-1/2}(2\varepsilon)^n \\
&\leq \varepsilon^n(\det\Sigma)^{-1/2}. \qquad (7)
\end{aligned}$$

where the covariance matrix

$$\Sigma = (\mathbb{E}Y(\delta_i)Y(\delta_j))_{1\leq i,j\leq n} = \left(\frac{1-e^{-\delta_i-\delta_j}}{\delta_i+\delta_j}\right)_{1\leq i,j\leq n}.$$

To find a lower bound for $\det(\Sigma)$, we need the following lemma:

Lemma 4.1. If $0 < b_{ij} < a_{ij}$ for all $1 \leq i, j \leq n$ then

$$\det(a_{ij} - b_{ij}) \geq \det(a_{ij}) - \sum_{k=1}^{n} \max_{1 \leq l \leq n} \frac{b_{kl}}{a_{kl}} \cdot \mathrm{per}(a_{ij}).$$

where $\mathrm{per}(a_{ij})$ is the permanent of the matrix (a_{ij}).

Proof. For notational simplicity, we denote $c_{ij} = a_{ij} - b_{ij}$, then

$$\begin{aligned}
& \det(a_{ij} - b_{ij}) - \det(a_{ij}) \\
= \; & \sum_{\sigma} (-1)^{\sigma} c_{1,\sigma(1)} c_{2,\sigma(2)} c_{n,\sigma(n)} - \sum_{\sigma} (-1)^{\sigma} a_{1,\sigma(1)} a_{2,\sigma(2)} \cdots a_{n,\sigma(n)} \\
= \; & \sum_{\sigma} (-1)^{\sigma} \sum_{k=1}^{n} [c_{1,\sigma(1)} \cdots c_{k-1,\sigma(k-1)}](c_{k,\sigma(k)} - a_{k,\sigma(k)})[a_{k+1,\sigma(k+1)} \cdots a_{n,\sigma(n)}] \\
\geq \; & -\sum_{\sigma} \sum_{k=1}^{n} [a_{1,\sigma(1)} \cdots a_{k-1,\sigma(k-1)}](b_{k,\sigma(k)})[a_{k+1,\sigma(k+1)} \cdots a_{n,\sigma(n)}] \\
\geq \; & -\sum_{k=1}^{n} \max_{1 \leq l \leq n} \frac{b_{kl}}{a_{kl}} \sum_{\sigma} [a_{1,\sigma(1)} \cdots a_{k-1,\sigma(k-1)}](a_{k,\sigma(k)})[a_{k+1,\sigma(k+1)} \cdots a_{n,\sigma(n)}] \\
= \; & -\sum_{k=1}^{n} \max_{1 \leq l \leq n} \frac{b_{kl}}{a_{kl}} \cdot \mathrm{per}(a_{ij}).
\end{aligned}$$

□

In order to use Lemma 4.1 to estimate $\det(\Sigma)$, we set

$$a_{ij} = \frac{1}{\delta_i + \delta_j}, \quad \text{and} \quad b_{ij} = e^{-\delta_i - \delta_j} a_{ij}$$

for a specific sequence $\{\delta_i\}_{i=1}^{n}$ defined by

$$\delta_{mp+q} = 4^{p+m}(m+q), \quad 0 \leq p < m, 1 \leq q \leq m$$

for $n = m^2$.

Clearly, we have

$$0 < b_{kl}/a_{kl} \leq e^{-2m4^m}, \quad 1 \leq k, l \leq n = m^2. \tag{8}$$

It remains to estimate $\det(a_{ij})$ and $\mathrm{per}(a_{ij})$, which is given in the following lemma.

Lemma 4.2. For the matrix (a_{ij}) defined above, we have $\mathrm{per}(a_{ij}) \leq 1$, and $\det(a_{ij}) \geq (240e)^{-2m^3}$.

Proof. It is easy to see

$$\mathrm{per}(a_{ij}) \leq n!(\max_{i,j} a_{ij})^n \leq \frac{(m^2)!}{(2m4^m)^{m^2}} \leq 1$$

since $a_{ij} \le (2m4^m)^{-1}$ for $1 \le i, j \le n = m^2$.

To estimate $\det(a_{ij})$, we use the Cauchy's determinant identity, see Krattenthaler (1999),

$$\det(a_{ij}) = \det\left(\frac{1}{\delta_i + \delta_j}\right) = \frac{\prod_{1\le i<j\le n}(\delta_j - \delta_i)^2}{\prod_{1\le i,j\le n}(\delta_j + \delta_i)} = \frac{1}{2^n \prod_{i=1}^n \delta_i} \cdot \prod_{1\le i<j\le n} \left(\frac{\delta_j - \delta_i}{\delta_j + \delta_i}\right)^2.$$

For $1 \le i < j \le n = m^2$, write $i = mp + q$ and $j = mr + s$ with $1 \le q, s \le m$. Denote

$$A = \{(i,j) : i = mp+q, j = mp+s, 0 \le p \le m-1, 1 \le q < s \le m\},$$
$$B = \{(i,j) : i = mp+q, j = m(p+1)+s, 0 \le p \le m-2, 1 \le q, s \le m\},$$
$$C = \{(i,j) : i = mp+q, j = mr+s, 0 \le p \le m-3, p+2 \le r \le m-1, 1 \le q, s \le m\}.$$

Then A, B and C form a partition of $1 \le i < j \le n = m^2$.

Next we estimate each part separately. First, for $p = r$,

$$\frac{\delta_j - \delta_i}{\delta_j + \delta_i} = \frac{s-q}{2m+s+q} > \frac{s-q}{4m}.$$

Thus

$$\begin{aligned}
\prod_{(i,j)\in A} \left(\frac{\delta_j - \delta_i}{\delta_j + \delta_i}\right)^2 &\ge \prod_{p=0}^{m-1} \prod_{1\le q<s\le m} \left(\frac{s-q}{4m}\right)^2 = \prod_{k=1}^{m-1}\prod_{q=1}^{m-k} \left(\frac{k}{4m}\right)^{2m} \\
&\ge \prod_{k=1}^{m-1} \left(\frac{k}{4m}\right)^{2m^2} = \left(\frac{(m-1)!}{(4m)^{m-1}}\right)^{2m^2} \\
&\ge (8e)^{-2m^3}.
\end{aligned}$$

Second, for $r - p = 1$,

$$\frac{\delta_j - \delta_i}{\delta_j + \delta_i} = \frac{(4m+4s) - (m+q)}{(4m+4s) + (m+q)} \ge \frac{1}{5}.$$

Thus we have

$$\prod_{(i,j)\in B} \left(\frac{\delta_j - \delta_i}{\delta_j + \delta_i}\right)^2 \ge \prod_{p=0}^{m-2} \prod_{1\le q,s\le m} 5^{-2} \ge 5^{-2m^3}.$$

Third, for $r - p \ge 2$,

$$\left(\frac{\delta_j - \delta_i}{\delta_j + \delta_i}\right) = \frac{4^r(m+s) - 4^p(m+q)}{4^r(m+s) + 4^p(m+q)} = 1 - \frac{2\cdot 4^p(m+q)}{4^r(m+s) + 4^p(m+q)} > 1 - \frac{1}{4^{r-p-1}}.$$

Thus we have

$$
\begin{aligned}
\prod_{(i,j)\in C}\left(\frac{\delta_j-\delta_i}{\delta_j+\delta_i}\right)^2 &\geq \prod_{p=0}^{m-3}\prod_{r=p+2}^{m-1}\prod_{1\leq q,s\leq m}\left(1-\frac{1}{4^{r-p-1}}\right)^2 \geq \prod_{k=1}^{m-2}\left(1-4^{-k}\right)^{2m^3} \\
&= \exp\left(-2m^3\sum_{k=1}^{m-2}\sum_{l=1}^{\infty}\frac{4^{-kl}}{l}\right) \\
&\geq \exp\left(-2m^3\sum_{l=1}^{\infty}\frac{1}{(4^l-1)l}\right) \geq \exp\left(-2m^3\sum_{l=1}^{\infty}\frac{1}{3^l l}\right) \\
&= (2/3)^{2m^3}.
\end{aligned}
$$

Therefore, we have

$$
\prod_{1\leq i<j\leq n}\left(\frac{\delta_j-\delta_i}{\delta_j+\delta_i}\right)^2 = \prod_{(i,j)\in A}\cdot\prod_{(i,j)\in B}\cdot\prod_{(i,j)\in C}\left(\frac{\delta_j-\delta_i}{\delta_j+\delta_i}\right)^2 \geq (60e)^{-2m^3}.
$$

On the other hand, it is not difficult to see that

$$
\begin{aligned}
2^n\prod_{i=1}^{n}\delta_i &= 2^{m^2}\prod_{q=1}^{m}\prod_{p=0}^{m-1}4^{p+m}(m+q) < 2^{m^2}\cdot 4^{m^2(m-1)/2+m^3}(2m)^{m^2} \\
&= 4^{3m^3/2+m^2/2+m^2\log_4 m} < 4^{2m^3}
\end{aligned}
$$

for $m>1$. Therefore,

$$
\det(a_{ij}) = \left(2^n\prod_{i=1}^{n}\delta_i\right)^{-1}\cdot\prod_{1\leq i<j\leq n}\left(\frac{\delta_j-\delta_i}{\delta_j+\delta_i}\right)^2 \geq (240e)^{-2m^3}.
$$

□

Now combining the two lemmas above, and using the estimate in (8), we obtain

$$
\det(\Sigma) \geq (240e)^{-2m^3} - m^2\cdot e^{-2m4^m} \geq e^{-16m^3}
$$

provided that m is large enough. Plugging into (7), we have

$$
\mathbb{P}\left(\sup_{t\geq 0}|Y(t)|<\varepsilon\right) \leq e^{8m^3}\varepsilon^{m^2}.
$$

Minimizing the right-hand side by choosing $m\approx|\log\varepsilon|/12$, we obtain

$$
\mathbb{P}\left(\sup_{t\geq 0}|Y(t)|<\varepsilon\right) \lesssim \exp\left(-(432)^{-1}|\log\varepsilon|^3\right).
$$

Statement (ii) of Theorem 1.1 follows by applying Corollary 3.2. At the same time, we also finished the proof of Theorem 1.2.

References

Alzer, H. and Berg, C. (2002). Some classes of completely monotonic functions. *Ann. Acad. Sci. Fenn. Math.* **27** 445–460.

Artstein, S., Milman, V., Szarek, S. and Tomczak-Jaegermann, N. (2004). On convexified packing and entropy duality. *Geom. Funct. Anal.* **14** 1134–1141.

Aurzada, F., Ibragimov, I., Lifshits, M. and van Zanten J.H. (2008). Approximation, metric entropy and small ball estimates for Gaussian measures. *preprint* .

Blei, R., Gao, F. and Li, W. V. (2007). Metric entropy of high dimensional distributions. *Proc. Amer. Math. Soc.* **135** 4009–4018.

Bourgain, J., Pajor, A., Szarek, S. J. and Tomczak-Jaegermann, N. (1989). On the duality problem for entropy numbers of operators. In *Geometric aspects of functional analysis (1987–88)*, vol. 1376 of *Lecture Notes in Math.* Springer, Berlin, 50–63.

Feller, W. (1971). *An introduction to probability theory and its applications. Vol. II.* Second edition, John Wiley & Sons Inc., New York.

Gao, F. (2004). Entropy of absolute convex hulls in Hilbert spaces. *Bull. London Math. Soc.* **36** 460–468.

Gao, F. (2008). Entropy estimate for k-monotone functions via small ball probability of integrated Brownian motion. *Electron. Commun. Probab.* **13** 121–130.

Gao, F. and Wellner, J. A. (2009). On the rate of convergence of the maximum likelihood estimator of a k-monotone density. *Science in China, Series A: Mathematics* **52** 1525–1538.

Jewell, N. P. (1982). Mixtures of exponential distributions. *Ann. Statist.* **10** 479–484.

Krattenthaler, C. (1999). Advanced determinant calculus. *Sém. Lothar. Combin.* **42** Art. B42q, 67 pp. (electronic). The Andrews Festschrift (Maratea, 1998).

Kuelbs, J. and Li, W. V. (1993). Metric entropy and the small ball problem for Gaussian measures. *J. Funct. Anal.* **116** 133–157.

Li, W. V. and Linde, W. (1999). Approximation, metric entropy and small ball estimates for Gaussian measures. *Ann. Probab.* **27** 1556–1578.

Tomczak-Jaegermann, N. (1987). Dualité des nombres d'entropie pour des opérateurs à valeurs dans un espace de Hilbert. *C. R. Acad. Sci. Paris Sér. I Math.* **305** 299–301.

Widder, D. V. (1941). *The Laplace Transform.* Princeton Mathematical Series, v. 6, Princeton University Press, Princeton, N. J.

Williamson, R. E. (1956). Multiply monotone functions and their Laplace transforms. *Duke Math. J.* **23** 189–207.